Published by IOP Publishing for Sissa Medialab

Received: *November 3, 2024*
Revised: *January 1, 2025*
Accepted: *February 11, 2025*
Published: *March 18, 2025*

Topical Workshop on Electronics for Particle Physics
University of Glasgow, Scotland, U.K.
30 September–4 October 2024

Firmware implementation of Phase-2 Overlap Muon Track Finder algorithm for CMS Level-1 trigger

Piotr Andrzej Fokow[a,*] **and Pelayo Leguina**[b] **on behalf of the CMS collaboration**

[a] *Institute of Electronic Systems, Warsaw University of Technology, Warsaw, Poland*
[b] *ICTEA, University of Oviedo, Oviedo, Spain*

E-mail: p.fokow@cern.ch

Abstract: The Overlap Muon Track Finder (OMTF) is one of the subsystems of the Compact Muon Solenoid (CMS) Level-1 Trigger. For the High-Luminosity Large Hadron Collider era (CMS phase-2 upgrade), a new version of the OMTF is currently under development. This upgraded version will be implemented on a custom ATCA board X2O, which houses a Xilinx UltraScale+ FPGA and 25 Gbps optical transceivers. This contribution focuses on the firmware implementation of the muon trigger algorithm and input data pre-processing, leveraging the High-Level Synthesis (HLS) technique. The paper presents current design and verification results, as well as experiences in using both standard and non-standard HLS development workflows.

Keywords: Trigger algorithms; Digital electronic circuits; Digital signal processing (DSP)

*Corresponding author.

https://doi.org/10.1088/1748-0221/20/03/C03022

Contents

1 Introduction

The Compact Muon Solenoid (CMS) detector will upgrade its muon detector system for its Phase-2 operation under HL-LHC. [1]. The amount of incoming data to the muon trigger algorithms will increase, and consequently, the algorithms must be optimized to handle it. This necessitates the implementation of a new version of the Overlap Muon Track Finder (OMTF) algorithm [2]. The OMTF algorithm was introduced for the Phase-1 CMS upgrade in 2016. It covers the pseudorapidity region $0.83 \leq |\eta| \leq 1.24$. The coverage of the OMTF algorithm is shown in figure 1. Its introduction was necessary because of specific detector geometry in this region and inequality in magnetic field. The OMTF algorithm for Phase-2, presented in figure 2, consists of several functional blocks: input interfaces, muon track reconstruction block, neural network, a ghostbuster module, and output

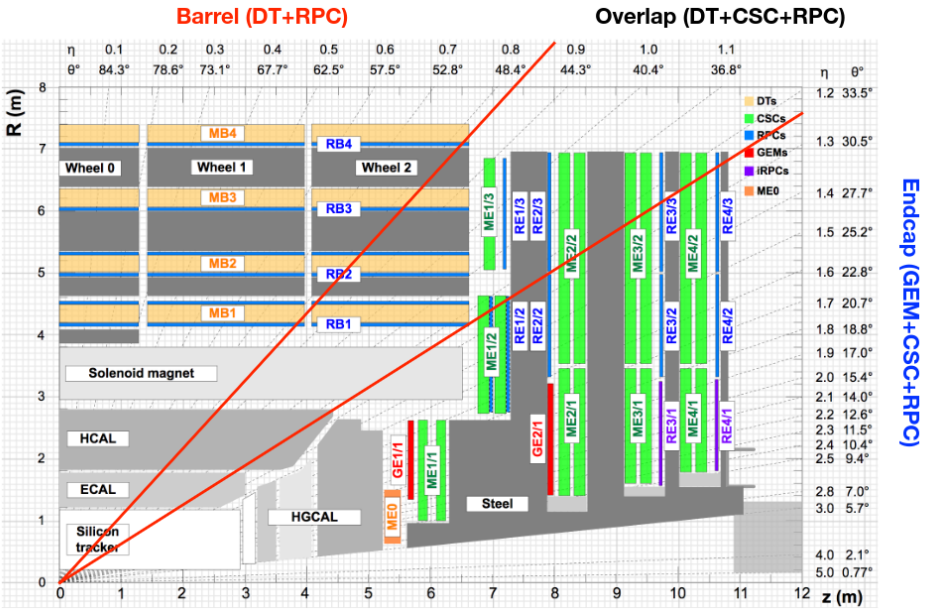

Figure 1. Longitudinal view of the quadrant of the CMS Phase-2 muon system with the OMTF algorithm range highlighted. Reproduced from [2]. CC BY 4.0.

2025 JINST 20 C03022

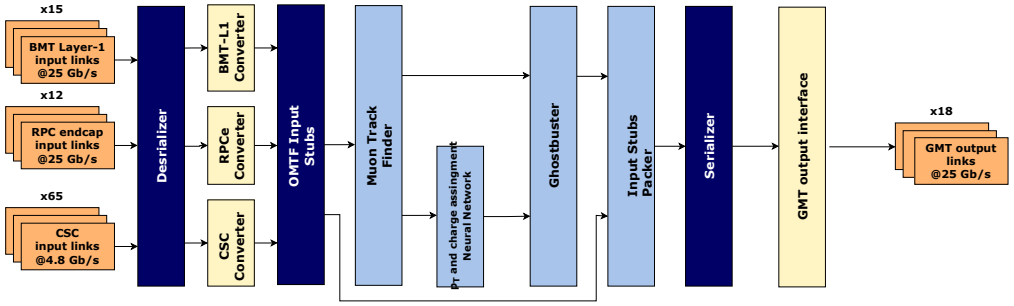

Figure 2. OMTF algorithm diagram for the Phase-2 CMS.

interfaces. The input interfaces take data from fiber optic links from the BMT-Layer 1 trigger system and Resistive Plate Chamber for Endcap (RPCe) and Cathode Strip Chamber (CSC) detectors. Then, they translate the data into the representation used in the OMTF algorithm. Input interfaces are also responsible for synchronizing the data. The muon track reconstruction algorithm builds muon tracks based on incoming muon hits [3]. The principle of muon reconstruction is shown in figure 3, where muon hits are presented as points, whereas real muon tracks are depicted as curved lines. Blue shapes represent the probability density functions (PDFs) for an individual pattern. The muon track is reconstructed, with hits from reference layers treated as the starting point. Only the hits with non-zero PDF values are considered for track reconstruction. The hits with the highest PDF values are selected for track matching. The analysis is performed in parallel for all patterns, and the pattern with the best matching parameter is selected as the muon candidate [4]. The neural network is introduced for Phase-2 and is further described in section 2. It calculates the estimated value of the transverse momentum and charge of the muon candidate. The muon candidates are then cleaned of duplicates in the ghostbuster block, after which, the candidates, together with the muon segments, are sent to the Global Muon Trigger (GMT) module.

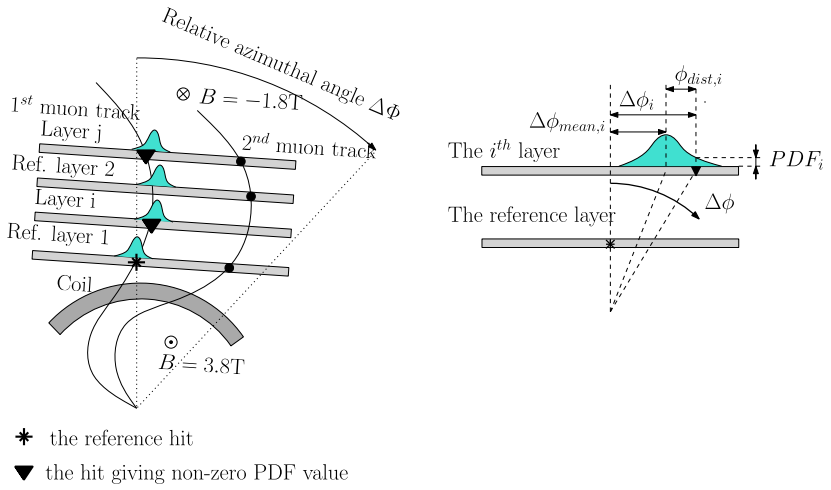

Figure 3. The principles of operation of Muon Track Reconstruction algorithm in OMTF. Reproduced from [4]. CC BY 4.0.

2025 JINST 20 C03022

2 Upgrades over Phase-1

The Phase-2 algorithm introduces new input and output interfaces. This paper focuses on the progress in implementing the BMT Layer-1 input interface, as the CSC, RPCe, and GMT interfaces are under development. The input interface block utilizes look-up tables (LUTs) to store the values required for translating the Phase-2 primitives to local coordinates, which are then used by further functional blocks.

The Muon Track Finder algorithm for Phase-2 is based on the algorithm from Run 3 [5]. There are some adjustments made for Phase-2. Patterns used for reconstruction and the reference hit selection algorithm have been adapted to the new trigger primitives. The pattern recognition algorithm was extended to enable triggering on muons produced by the decays of long-lived particles. This was achieved by measuring the transverse momentum without relying on the beam spot constraint. An extrapolated ϕ coordinate was introduced, which is calculated under the assumption that the data come from high-momentum tracks resulting from displaced decays [6]. This approach reduces the probability of reconstructing such tracks as low-momentum ones originating from beam interaction. In the implementation, the ϕ extrapolation factors are stored in LUTs. The neural network, as mentioned before, is introduced in Phase-2 OMTF algorithm. It is built in a fully-connected configuration [7]. The diagram of the neural network configuration and the implementation of a single neuron is presented in figure 4. The neural network input accepts the azimuthal angle difference relative to the reference hit found by the pattern logic ($\Delta\phi$) in every OMTF layer for every muon candidate. The algorithm consists of four layers in total, with two hidden layers. The output layer comprises two neurons, one responsible for the calculation of p_T and the other for the calculation of q. The neural network uses an optimization technique. In each neuron, values are taken from adjacent LUT positions, from which the difference is calculated. The input value is then multiplied by the difference obtained from the LUT, and finally, the previously retrieved table value is added to the input value. As a result, the calculation and activation function is performed in a single DSP48 block. This reduces the use of LUTs in the area of a single neuron, lowering the use of resources and the delay introduced by the neural network algorithm. The OMTF algorithm is implemented on a new hardware platform. It uses a backend system utilizing Virtex Ultrascale+ FPGA manufactured by AMD. The system is provided and developed by the CMS collaboration. For Phase-2, the overlap area will be handled by six trigger cards, each covering 120 degrees of the ϕ area with an overlap of 30 degrees. The entire

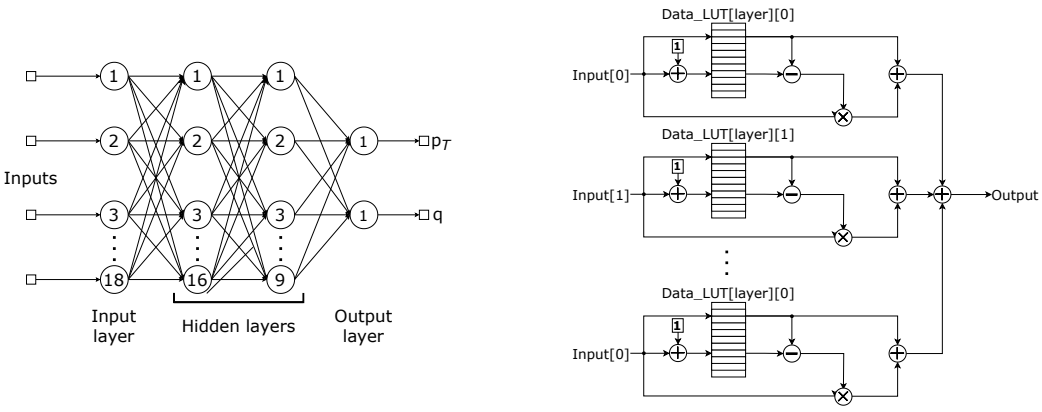

Figure 4. Neural network configuration (on the left) and neuron implementation (on the right).

algorithm operates at 360 MHz, compared to the Phase-1 algorithm, which operates at 160 MHz [8]. This increases the maximum output rate of the algorithm from 4 to 9 muons per bunch-crossing.[1] This is required due to the expected higher probability of generating new data.

3 Design and verification procedure

Along with the OMTF algorithm for Phase-2, a design and verification procedure for implementing the algorithm is also being developed (see figure 5). The procedure starts with the design of the IP cores. These are prepared based on the code used in CMSSW. The component IP blocks implemented with High-Level Synthesis (HLS) or HDL languages are verified separately against Monte Carlo simulation data from CMSSW. This ensures agreement between CMSSW and implementation. The IP cores are then synthesized and integrated within the Vivado design suite to produce the payload. The integrated design is once again verified. The design then goes through the implementation process. In this stage, functional block placement constraints, transceiver allocation, and optimization configurations are applied. In the future, the tests will be carried out on the hardware system, and the in-hardware verification step will be included in the procedure. The process is automated using CMake and TCL scripts so that the steps from testing to generating the binary bitstream for the FPGA device can be done using shell commands.

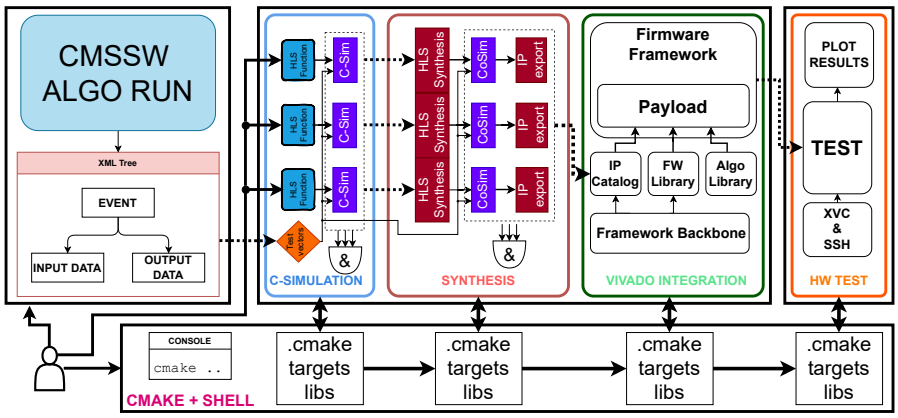

Figure 5. Design and verification procedure for OMTF algorithm.

4 Implementation results

The target device for the implementation is AMD UltraScale+ xcvu13p-fsga2577-1-e. In its current configuration, as shown in figure 6, the OMTF algorithm processes reduced inputs, specifically trigger primitives from BMT-Layer 1, for muon reconstruction. The implementation results are presented in table 1.

The implementation in its current state shows low resource usage, allowing it to fit into a single SLR block. The relatively high Block RAM (BRAM) utilization is due to the intensive use of BRAM LUTs to store neural network coefficients, Muon Track Finder algorithm patterns, and data for calculating input processing blocks. It should be noted that the use of LUT and Flip-Flop

[1]The bunch-crossing period is 25 ns.

2025 JINST 20 C03022

Table 1. Implementation results of OMTF algorithm for the system clock frequency of 360 MHz.

Parameter	Value
LUT utilization	6.3%
Flip-Flop utilization	5.3%
BRAM utilization	20.5%
DSP48 utilization	4.3%
Latency	328 ns
Worst Negative Slack	0.03 ns

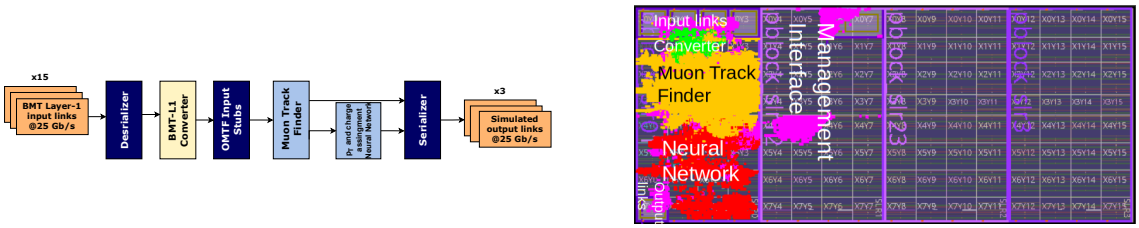

Figure 6. Diagram of test configuration of OMTF algorithm and implementation's floorplan.

resources will increase with the implementation of Endcap block data inputs. At this point, unused features are optimized out by behavioral synthesis and implementation. Verification was performed in two stages. First, all functional blocks were verified using long test vectors generated in CMSSW. Then, the integrated algorithm was run through a short data vector test to confirm the correct data propagation. The results show that the IP blocks are synthesized successfully, and the integrated algorithm propagates the data correctly. When increasing the clock frequency of the algorithm from 160 MHz for Phase-1 CMS to 360 MHz for Phase-2 CMS, some problems were encountered with read-write assumptions performed in a single clock beat. These were resolved by refactoring the code.

5 Conclusion and further development

The results showed that the OMTF algorithm was synthesized and the data propagation was executed correctly. Certain areas of the algorithm implementation need to be further explored, such as the FPGA algorithm floorplan. In its current state, the algorithm has not achieved its full functionality. A Ghostbuster module, which will clean the muon candidates from duplicates, is still being implemented. Additionally, the input interfaces from the detectors on the endcap side are being rebuilt. The output interface will be introduced to send data to the Global Muon Trigger module. Among the new features, parameter calculation without beam spot constraint will be integrated in the neural network.

Acknowledgments

The work was published thanks to the grant provided by the National Science Centre of Poland within the 2021/43/B/ST2/01552 project. The authors would also like to thank Karol Buńkowski for helping us coordinate with other CMS participants and providing us with deep knowledge about the CMS experiment.

2025 JINST 20 C03022

References

[1] CMS collaboration, *The CMS Experiment at the CERN LHC*, 2008 *JINST* **3** S08004.

[2] CMS collaboration, *The Phase-2 Upgrade of the CMS Level-1 Trigger*, CERN-LHCC-2020-004, CMS-TDR-021, CERN, Geneva (2020).

[3] CMS collaboration, *The algorithm of the CMS Level-1 Overlap Muon Track Finder trigger*, *Nucl. Instrum. Meth. A* **936** (2019) 368.

[4] M. Bluj et al., *From the physical model to the electronic system — OMTF Trigger for CMS*, *Acta Phys. Polon. Supp.* **9** (2016) 181.

[5] W.M. Zabołotny, *Implementation of OMTF trigger algorithm with high-level synthesis*, *Proc. SPIE Int. Soc. Opt. Eng.* **11176** (2019) 1117641.

[6] P. Leguina, *Firmware implementation of a displaced muon reconstruction algorithm for the Phase-2 Upgrade of the CMS muon system*, 2023 *JINST* **18** C12005.

[7] P. Petersen and F. Voigtlaender, *Equivalence of approximation by convolutional neural networks and fully-connected networks*, arXiv:1809.00973.

[8] W.M. Zabołotny and A.P. Byszuk, *Algorithm and implementation of muon trigger and data transmission system for barrel-endcap overlap region of the CMS detector*, 2016 *JINST* **11** C03004.

2025 JINST 20 C03022